LP Decoding meets LP Decoding: A Connection between Channel Coding and Compressed Sensing*

Alexandros G. Dimakis and Pascal O. Vontobel

***Abstract*— This is a tale of two linear programming decoders, namely channel coding linear programming decoding (CC-LPD) and compressed sensing linear programming decoding (CS-LPD). So far, they have evolved quite independently. The aim of the present paper is to show that there is a tight connection between, on the one hand, CS-LPD based on a zero-one measurement matrix over the reals and, on the other hand, CC-LPD of the binary linear code that is obtained by viewing this measurement matrix as a binary parity-check matrix. This connection allows one to translate performance guarantees from one setup to the other.**

I. Introduction

Recently there has been substantial interest in the theory of recovering sparse approximations of signals that satisfy linear measurements. Compressed (or compressive) sensing research (see, e.g., [1], [2]) has developed conditions for measurement matrices under which (approximately) sparse signals can be recovered by solving a linear programming relaxation of the original NP-hard combinatorial problem. Interestingly, in one of the first papers in this area (cf. [1]), Candes and Tao presented a setup they called "decoding by linear programming," henceforth called **CS-LPD**, where the sparse signal corresponds to real-valued noise that is added to a real-valued signal that is to be recovered in a hypothetical communication problem.

At about the same time, in an independent line of research, Feldman, Wainwright, and Karger considered the problem of decoding a binary linear code that is used for data communication over a binary-input memoryless channel, a problem that is also NP-hard in general. In [3], [4], they formulated this channel coding problem as an integer linear program, along with presenting a linear programming relaxation for it, henceforth called **CC-LPD**. Several theoretical results were subsequently proven about the efficiency of **CC-LPD**, in particular for low-density parity-check (LDPC) codes (e.g. [5], [6], [7], [8]).

As we will see in the subsequent sections, **CS-LPD** and **CC-LPD** (and the setups they are derived from) are *formally* very similar, however, it is rather unclear if there is a connection beyond this formal relationship. In fact Candes and Tao in their original paper asked the following question [1, Section VI.A]: "*. . . In summary, there does not seem to be any explicit known connection with this line of work*[1] *but it would perhaps be of future interest to explore if there is one.*"

In this paper we present such a connection between **CS-LPD** and **CC-LPD**. The general form of our results is that if a given binary parity-check matrix is "good" for **CC-LPD** then the same matrix (considered over the reals) is a "good" measurement matrix for **CS-LPD**. The notion of a "good" parity-check matrix depends on which channel we use (and a corresponding channel-dependent quantity called pseudo-weight).

- Based on results for the binary symmetric channel (BSC), we show that if a parity-check matrix can correct any k bit-flipping errors under **CC-LPD**, then the same matrix taken as a measurement matrix over the reals can be used to recover all k-sparse error signals under **CS-LPD**.
- Based on results for binary-input output-symmetric channels with bounded log-likelihood ratios, we can extend the previous result to show that performance guarantees for **CC-LPD** for such channels can be translated into robust sparse-recovery guarantees in the ℓ_1/ℓ_1 sense (see, e.g., [9]) for **CS-LPD**.
- Performance guarantees for **CC-LPD** for the binary-input AWGNC (additive white Gaussian noise channel) can be translated into robust sparse-recovery guarantees in the ℓ_2/ℓ_1 sense for **CS-LPD**
- Max-fractional weight performance guarantees for **CC-LPD** can be translated into robust sparse-recovery guarantees in the ℓ_∞/ℓ_1 sense for **CS-LPD**.
- Performance guarantees for **CC-LPD** for the BEC (binary erasure channel) can be translated into performance guarantees for the compressed sensing setup where the support of the error signal is known and the decoder tries to recover the sparse signal (i.e., tries to solve the linear equations) by back-substitution only.

All our results are also valid in a stronger, point-wise sense. For example, for the BSC, if a parity-check matrix can recover a *given set* of k bit flips under **CC-LPD**, the same matrix will recover any sparse signal supported on those k coordinates under **CS-LPD**. In general, "good" performance of **CC-LPD** on a given error support will yield "good" **CS-LPD** recovery for sparse signals supported on the same support.

It should be noted that all our results are only one-way: we

A. G. Dimakis is with the Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089, USA (email: dimakis@usc.edu).

P. O. Vontobel is with Hewlett–Packard Laboratories, 1501 Page Mill Road, Palo Alto, CA 94304, USA (email: pascal.vontobel@ieee.org).

*Appeared in the Proceedings of the 47th Allerton Conference on Communications, Control, and Computing, Allerton House, Monticello, Illinois, USA, Sep. 30–Oct. 2, 2009. This version of the paper contains all the proofs that were omitted in the official version due to space limitations.

[1] Candes and Tao [1, Section VI.A] refer here to [3], [4].

do not prove that a "good" zero-one measurement matrix will always be a "good" parity-check matrix for a binary code. This remains an interesting open problem.

The remainder of this paper is organized as follows. In Section II we set up the notation that will be used. Then in Sections III and IV we will review the compressed sensing and channel coding setups that we are interested in, along with their respective linear programming relaxations. This review will be presented in such a way that the close *formal* relationship between the two setups will stand out. Afterwards, in Section V we will show that for a zero-one matrix, once seen as a real-valued measurement matrix, once seen as a binary parity-check matrix, this close relationship is *not only formal* but that in fact non-zero vectors in the real nullspace of this matrix (i.e., vectors that are problematic vectors for **CS-LPD**) can be mapped to non-zero vectors in the fundamental cone defined by that same matrix (i.e., to vectors that are problematic vectors for **CC-LPD**). Based on this observation one can, as will be shown in Section VI, translate performance guarantees from one setup to the other. The paper finishes with some conclusions in Section VII.

II. Basic Notation

Let $\mathbb{Z}$, $\mathbb{Z}_{\geqslant 0}$, $\mathbb{Z}_{>0}$, $\mathbb{R}$, $\mathbb{R}_{\geqslant 0}$, $\mathbb{R}_{>0}$, and $\mathbb{F}_2$ be the ring of integers, the set of non-negative integers, the set of positive integers, the field of real numbers, the set of non-negative real numbers, the set of positive real numbers, and the finite field of size 2, respectively. Unless noted otherwise, expressions, equalities, and inequalities will be over the field $\mathbb{R}$. The absolute value of a real number a will be denoted by $|a|$. The size of a set $\mathcal{S}$ will be denoted by $\#\mathcal{S}$.

In this paper all vectors will be *column* vectors. If $\boldsymbol{a}$ is some vector with integer entries, then $\boldsymbol{a} \pmod 2$ will denote an equally long vector whose entries are reduced modulo 2. If $\mathcal{S}$ is a subset of the set of coordinate indices of a vector $\boldsymbol{a}$ then $\boldsymbol{a}_{\mathcal{S}}$ is the vector of length $\#\mathcal{S}$ that contains only the coordinates of $\boldsymbol{a}$ whose coordinate index appears in $\mathcal{S}$. Moreover, if $\boldsymbol{a}$ is a real vector then we define $|\boldsymbol{a}|$ to be the real vector $\boldsymbol{a}'$ of the same length as $\boldsymbol{a}$ with entries $a'_i = |a_i|$ for all i. Finally, the inner product $\langle \boldsymbol{a}, \boldsymbol{b} \rangle$ of two equally long vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is defined to $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \sum_i a_i b_i$.

We define $\operatorname{supp}(\boldsymbol{a}) \triangleq \{i \mid a_i \neq 0\}$ to be the support set of some vector $\boldsymbol{a}$. Moreover, we let $\Sigma^{(k)}_{\mathbb{R}^n} \triangleq \big\{\boldsymbol{a} \in \mathbb{R}^n \;\big|\; \#\operatorname{supp}(\boldsymbol{a}) \leqslant k\big\}$ and $\Sigma^{(k)}_{\mathbb{F}_2^n} \triangleq \big\{\boldsymbol{a} \in \mathbb{F}_2^n \;\big|\; \#\operatorname{supp}(\boldsymbol{a}) \leqslant k\big\}$ be the set of vectors in $\mathbb{R}^n$ and $\mathbb{F}_2^n$, respectively, which have at most k non-zero components. If $k \ll n$ then vectors in these sets are called k-sparse vectors.

For any real vector $\boldsymbol{a}$, we define $\|\boldsymbol{a}\|_0$ to be the ℓ_0 norm of $\boldsymbol{a}$, i.e., the number of non-zero components of $\boldsymbol{a}$. Note that $\|\boldsymbol{a}\|_0 = w_{\mathrm{H}}(\boldsymbol{a}) = |\operatorname{supp}(\boldsymbol{a})|$, where $w_{\mathrm{H}}(\boldsymbol{a})$ is the Hamming weight of $\boldsymbol{a}$. Furthermore, $\|\boldsymbol{a}\|_1 \triangleq \sum_i |a_i|$, $\|\boldsymbol{a}\|_2 \triangleq \sqrt{\sum_i |a_i|^2}$, $\|\boldsymbol{a}\|_\infty \triangleq \max_i |a_i|$ will denote, respectively, the ℓ_1, ℓ_2, and ℓ_∞ norm of $\boldsymbol{a}$.

For a matrix $\boldsymbol{M}$ over $\mathbb{R}$ with n columns we define its $\mathbb{R}$ nullspace to be the set $\operatorname{nullspace}_{\mathbb{R}}(\boldsymbol{H}) \triangleq \big\{\boldsymbol{a} \in \mathbb{R}^n \;\big|\; \boldsymbol{M} \cdot \boldsymbol{a} = \boldsymbol{0}\big\}$ and for a matrix $\boldsymbol{M}$ over $\mathbb{F}_2$ with n columns we define its $\mathbb{F}_2$ nullspace to be the set $\operatorname{nullspace}_{\mathbb{F}_2}(\boldsymbol{H}) \triangleq \big\{\boldsymbol{a} \in \mathbb{F}_2^n \;\big|\; \boldsymbol{M} \cdot \boldsymbol{a} = \boldsymbol{0} \text{ (in } \mathbb{F}_2)\big\}$.

Let $\boldsymbol{H} = (h_{j,i})_{j,i}$ be some matrix. We define the sets $\mathcal{J}(\boldsymbol{H})$ and $\mathcal{I}(\boldsymbol{H})$ to be, respectively, the set of row and column indices of $\boldsymbol{H}$. Moreover, we will use the sets $\mathcal{J}_i(\boldsymbol{H}) \triangleq \{j \in \mathcal{J} \mid h_{j,i} \neq 0\}$ and $\mathcal{I}_j(\boldsymbol{H}) \triangleq \{i \in \mathcal{I} \mid h_{j,i} \neq 0\}$. In the following, when no confusion can arise, we will sometimes omit the argument $\boldsymbol{H}$ in the preceding expressions. For any set $\mathcal{S} \subseteq \mathcal{I}$, we will denote its complement with respect to $\mathcal{I}$ by $\overline{\mathcal{S}}$, i.e., $\overline{\mathcal{S}} \triangleq \mathcal{I} \setminus \mathcal{S}$.

III. Compressed Sensing Linear Programming Decoding

A. The Setup

Let $\boldsymbol{H}_{\mathrm{CS}}$ be a real matrix of size $m \times n$, called the measurement matrix, and let $\boldsymbol{s}$ be a real vector of length m. In its simplest form, the compressed sensing problem consists of finding the sparsest real vector $\boldsymbol{e}'$ of length n that satisfies $\boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{e}' = \boldsymbol{s}$, namely

$$\boxed{\begin{array}{lll} \textbf{CS-OPT}: & \text{minimize} & \|\boldsymbol{e}'\|_0 \\ & \text{subject to} & \boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{e}' = \boldsymbol{s}. \end{array}}$$

Assuming that there exists a truly sparse signal $\boldsymbol{e}$ that satisfies the measurement $\boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{e} = \boldsymbol{s}$, **CS-OPT** yields, for suitable matrices $\boldsymbol{H}_{\mathrm{CS}}$, an estimate $\hat{\boldsymbol{e}}$ that equals $\boldsymbol{e}$.

This problem can also be interpreted [1] as part of the decoding problem that appears in a coded data communicating setup where the channel input alphabet is $\mathcal{X}_{\mathrm{CS}} \triangleq \mathbb{R}$, the channel output alphabet is $\mathcal{Y}_{\mathrm{CS}} \triangleq \mathbb{R}$, and the information symbols are encoded with the help of a real-valued code $\mathcal{C}_{\mathrm{CS}}$ of length n and dimension $\kappa \triangleq n - \operatorname{rank}_{\mathbb{R}}(\boldsymbol{H}_{\mathrm{CS}})$ as follows.

- The code is $\mathcal{C}_{\mathrm{CS}} \triangleq \big\{\boldsymbol{x} \in \mathbb{R}^n \;\big|\; \boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{x} = \boldsymbol{0}\big\}$. Because of this, the measurement matrix $\boldsymbol{H}_{\mathrm{CS}}$ is sometimes also called an annihilator matrix.
- A matrix $\boldsymbol{G}_{\mathrm{CS}} \in \mathbb{R}^{n \times \kappa}$ for which $\mathcal{C}_{\mathrm{CS}} = \big\{\boldsymbol{G}_{\mathrm{CS}} \cdot \boldsymbol{u} \;\big|\; \boldsymbol{u} \in \mathbb{R}^\kappa\big\}$ holds, is called a generator matrix for the code $\mathcal{C}_{\mathrm{CS}}$. With the help of such a matrix, information vectors $\boldsymbol{u} \in \mathbb{R}^\kappa$ are encoded into codewords $\boldsymbol{x} \in \mathbb{R}^n$ according to $\boldsymbol{x} = \boldsymbol{G}_{\mathrm{CS}} \cdot \boldsymbol{u}$.
- Let $\boldsymbol{y} \in \mathcal{Y}_{\mathrm{CS}}^n$ be the received vector. We can write $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}$ for a suitably defined vector $\boldsymbol{e} \in \mathbb{R}^n$, which will be called the error vector. We assume that the channel is such that $\boldsymbol{e}$ is sparse or approximately sparse.
- The receiver first computes the syndrome vector $\boldsymbol{s}$ according to $\boldsymbol{s} \triangleq \boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{y}$. Note that

  $$\begin{aligned} \boldsymbol{s} &= \boldsymbol{H}_{\mathrm{CS}} \cdot (\boldsymbol{x} + \boldsymbol{e}) = \boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{x} + \boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{e} \\ &= \boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{e}. \end{aligned}$$

  In a second step, the receiver solves **CS-OPT** to obtain an estimate $\hat{\boldsymbol{e}}$ for $\boldsymbol{e}$, which can be used to obtain the codeword estimate $\hat{\boldsymbol{x}} = \boldsymbol{y} - \hat{\boldsymbol{e}}$, which in turn can be used to obtain the information word estimate $\hat{\boldsymbol{u}}$.

Because the complexity of solving **CS-OPT** is usually exponential in the relevant parameters, one can try to formulate and solve a related optimization problem with the aim that the related optimization problem yields very often the same solution as **CS-OPT**, or at least very often a very good approximation to the solution given by **CS-OPT**. In the context of **CS-OPT**, a popular approach is to formulate and solve the following related optimization problem (which, with the suitable introduction of auxiliary variables, can be turned into a linear program):

$$\boxed{\begin{array}{lll} \textbf{CS-LPD}: & \text{minimize} & \|\boldsymbol{e}'\|_1 \\ & \text{subject to} & \boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{e}' = \boldsymbol{s}. \end{array}}$$

B. Conditions for the Equivalence of ***CS-LPD*** and ***CS-OPT***

A central question of compressed sensing theory is under what conditions the solution given by **CS-LPD** equals (or is very close to) the solution given by **CS-OPT**.[2] Clearly, if $m \geqslant n$ and the matrix $\boldsymbol{H}_{\mathrm{CS}}$ has rank n, there is only one feasible $\boldsymbol{e}'$ and the two problems have the same solution.

In this paper we typically focus on the linear sparsity regime, i.e., $k = \Theta(n)$ and $m = \Theta(n)$, but our techniques are more generally applicable. The question is for which measurement matrices (hopefully with a small number of measurements m) the LP relaxation is tight, i.e., the estimate given by **CS-LPD** equals the estimate given by **CS-OPT**. One such *sufficient* condition is that a given measurement matrix is "good" if it satisfies the restricted isometry property (RIP), i.e., does not distort the ℓ_2 length of all k-sparse vectors. If this is the case then it was shown [1] that the LP relaxation will be tight for all k-sparse vectors $\boldsymbol{e}$ and further the recovery will be robust to approximate sparsity. The RIP condition however is not a complete characterization of "good" measurement matrices. We will use the nullspace characterization (see, e.g., [10], [11]) instead, that is necessary and sufficient.

Definition 1 *Let $\mathcal{S} \subseteq \mathcal{I}(\boldsymbol{H}_{\mathrm{CS}})$ and let $C \in \mathbb{R}_{\geqslant 0}$. We say that $\boldsymbol{H}_{\mathrm{CS}}$ has the nullspace property $\mathrm{NSP}^{\leqslant}_{\mathbb{R}}(\mathcal{S}, C)$, and write $\boldsymbol{H}_{\mathrm{CS}} \in \mathrm{NSP}^{\leqslant}_{\mathbb{R}}(\mathcal{S}, C)$, if*

$$C \cdot \|\boldsymbol{\nu}_{\mathcal{S}}\|_1 \leqslant \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 \quad \textit{for all} \quad \boldsymbol{\nu} \in \mathrm{nullspace}_{\mathbb{R}}(\boldsymbol{H}_{\mathrm{CS}}).$$

We say that $\boldsymbol{H}_{\mathrm{CS}}$ has the strict nullspace property $\mathrm{NSP}^{<}_{\mathbb{R}}(\mathcal{S}, C)$, and write $\boldsymbol{H}_{\mathrm{CS}} \in \mathrm{NSP}^{<}_{\mathbb{R}}(\mathcal{S}, C)$, if

$$C \cdot \|\boldsymbol{\nu}_{\mathcal{S}}\|_1 < \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 \quad \textit{for all} \quad \boldsymbol{\nu} \in \mathrm{nullspace}_{\mathbb{R}}(\boldsymbol{H}_{\mathrm{CS}}) \setminus \{\boldsymbol{0}\}.$$

□

[2]It is important to note that we worry only about the solution given by **CS-LPD** being equal (or very close to) the solution given by **CS-OPT**, because even **CS-OPT** might fail to correctly estimate the error vector in the above communication setup when the error vector has too many large components.

Definition 2 *Let $k \in \mathbb{Z}_{\geqslant 0}$ and let $C \in \mathbb{R}_{\geqslant 0}$. We say that $\boldsymbol{H}_{\mathrm{CS}}$ has the nullspace property $\mathrm{NSP}^{\leqslant}_{\mathbb{R}}(k, C)$, and write $\boldsymbol{H}_{\mathrm{CS}} \in \mathrm{NSP}^{\leqslant}_{\mathbb{R}}(k, C)$, if*

$$\boldsymbol{H}_{\mathrm{CS}} \in \mathrm{NSP}^{\leqslant}_{\mathbb{R}}(\mathcal{S}, C) \quad \textit{for all} \quad \mathcal{S} \subseteq \mathcal{I}(\boldsymbol{H}_{\mathrm{CS}}) \textit{ with } \#\mathcal{S} \leqslant k.$$

We say that $\boldsymbol{H}_{\mathrm{CS}}$ has the strict nullspace property $\mathrm{NSP}^{<}_{\mathbb{R}}(k, C)$, and write $\boldsymbol{H}_{\mathrm{CS}} \in \mathrm{NSP}^{<}_{\mathbb{R}}(k, C)$, if

$$\boldsymbol{H}_{\mathrm{CS}} \in \mathrm{NSP}^{<}_{\mathbb{R}}(\mathcal{S}, C) \quad \textit{for all} \quad \mathcal{S} \subseteq \mathcal{I}(\boldsymbol{H}_{\mathrm{CS}}) \textit{ with } \#\mathcal{S} \leqslant k.$$

□

As was shown independently by several authors (see [12], [13], [14], [11] and references therein) the nullspace condition in Definition 2 is a necessary and sufficient condition for a measurement matrix to be "good" for k-sparse signals, i.e. that the estimate given by **CS-LPD** equals the estimate given by **CS-OPT** for these matrices. The nullspace characterization of "good" measurement matrices will be one of the keys to linking **CS-LPD** with **CC-LPD**. Observe that the requirement is that vectors in the nullspace of $\boldsymbol{H}_{\mathrm{CS}}$ have their ℓ_1 mass spread in substantially more than k coordinates. The following theorem is adapted from [11] (and references therein).

Theorem 3 *Let $\boldsymbol{H}_{\mathrm{CS}}$ be a measurement matrix. Further, assume that $\boldsymbol{s} = \boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{e}$ and that $\boldsymbol{e}$ has at most k nonzero elements, i.e., $\|\boldsymbol{e}\|_0 \leqslant k$. Then the estimate $\hat{\boldsymbol{e}}$ produced by **CS-LPD** will equal the estimate $\hat{\boldsymbol{e}}$ produced by **CS-OPT** if $\boldsymbol{H}_{\mathrm{CS}} \in \mathrm{NSP}^{<}_{\mathbb{R}}(k, C = 1)$.*

Remark: Actually, as discussed in [11] and references therein, the condition $\boldsymbol{H}_{\mathrm{CS}} \in \mathrm{NSP}^{<}_{\mathbb{R}}(k, C = 1)$ is also necessary, but we will not use this here.

The next performance metric (see, e.g., [9], [15]) for CS involves recovering sparse approximations to signals that are not exactly k-sparse.

Definition 4 *An ℓ_p/ℓ_q approximation guarantee for **CS-LPD** means that the **CS-LPD** outputs an estimate $\hat{\boldsymbol{e}}$ that is within a factor $C_{p,q}(k)$ from the best k-sparse approximation for $\boldsymbol{e}$, i.e.,*

$$\|\boldsymbol{e} - \hat{\boldsymbol{e}}\|_p \leqslant C_{p,q}(k) \cdot \min_{\boldsymbol{e}' \in \Sigma^{(k)}_{\mathbb{R}^n}} \|\boldsymbol{e} - \boldsymbol{e}'\|_q, \tag{1}$$

where the left-hand side is measured in the ℓ_p norm and the right-hand side is measured in the ℓ_q norm. □

Note that the minimizer of the right-hand side of (1) (for any norm) is the vector $\boldsymbol{e}' \in \Sigma^{(k)}_{\mathbb{R}^n}$ that has the k largest (in magnitude) coordinates of $\boldsymbol{e}$, also called the best k-term approximation of $\boldsymbol{e}$ [15]. Therefore the right-hand side of (1) equals $C_{p,q}(k) \cdot \|\boldsymbol{e}_{\overline{\mathcal{S}^*}}\|_q$ where $\mathcal{S}^*$ is the support set of the k largest (in magnitude) components of $\boldsymbol{e}$. Also note that if $\boldsymbol{e}$ is exactly k-sparse the above condition suggests that $\hat{\boldsymbol{e}} = \boldsymbol{e}$ since the right hand-side of (1) vanishes, therefore it is a strictly stronger statement than recovery of sparse signals. (Of course, such a stronger approximation guarantee for $\hat{\boldsymbol{e}}$ is usually only obtained under stronger assumptions on the measurement matrix.)

The nullspace condition is necessary and sufficient for ℓ_1/ℓ_1 approximation for any measurement matrix. This is shown in the next theorem and proof which are adapted from [10, Theorem 1]. (Actually, we omit the necessity part in the next theorem since it will not be needed in this paper.)

Theorem 5 *Let $\boldsymbol{H}_{\mathrm{CS}}$ be a measurement matrix and choose some constant $C > 1$. Further, assume that $\boldsymbol{s} = \boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{e}$. Then for any set $\mathcal{S} \subseteq \mathcal{I}$ with $\#\mathcal{S} \leqslant k$ the solution $\hat{\boldsymbol{e}}$ produced by* ***CS-LPD*** *will satisfy*

$$\|\boldsymbol{e} - \hat{\boldsymbol{e}}\|_1 \leqslant 2 \cdot \frac{C+1}{C-1} \cdot \|\boldsymbol{e}_{\overline{\mathcal{S}}}\|_1$$

if $\boldsymbol{H}_{\mathrm{CS}} \in \mathrm{NSP}^{\leqslant}_{\mathbb{R}}(k, C)$.

Proof: Suppose that $\boldsymbol{H}_{\mathrm{CS}}$ has the claimed nullspace property. Since $\boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{e} = \boldsymbol{s}$ and $\boldsymbol{H}_{\mathrm{CS}} \cdot \hat{\boldsymbol{e}} = \boldsymbol{s}$, it easily follows that $\boldsymbol{\nu} \triangleq \boldsymbol{e} - \hat{\boldsymbol{e}}$ is in the nullspace of $\boldsymbol{H}_{\mathrm{CS}}$. So,

$$\begin{aligned}
\|\boldsymbol{e}_{\mathcal{S}}\|_1 + \|\boldsymbol{e}_{\overline{\mathcal{S}}}\|_1 &= \|\boldsymbol{e}\|_1 \\
&\overset{\text{(a)}}{\geqslant} \|\hat{\boldsymbol{e}}\|_1 \\
&= \|\boldsymbol{e} + \boldsymbol{\nu}\|_1 \\
&= \|\boldsymbol{e}_{\mathcal{S}} + \boldsymbol{\nu}_{\mathcal{S}}\|_1 + \|\boldsymbol{e}_{\overline{\mathcal{S}}} + \boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 \\
&\overset{\text{(b)}}{\geqslant} \|\boldsymbol{e}_{\mathcal{S}}\|_1 - \|\boldsymbol{\nu}_{\mathcal{S}}\|_1 + \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 - \|\boldsymbol{e}_{\overline{\mathcal{S}}}\|_1 \\
&\overset{\text{(c)}}{\geqslant} \|\boldsymbol{e}_{\mathcal{S}}\|_1 + \frac{C-1}{C+1} \cdot \|\boldsymbol{\nu}\|_1 - \|\boldsymbol{e}_{\overline{\mathcal{S}}}\|_1, \qquad (2)
\end{aligned}$$

where step (a) follows from the fact that the solution to **CS-LPD** satisfies $\|\hat{\boldsymbol{e}}\|_1 \leqslant \|\boldsymbol{e}\|_1$, where step (b) follows from applying the triangle inequality for the ℓ_1 norm twice, and where step (c) follows from

$$-\|\boldsymbol{\nu}_{\mathcal{S}}\|_1 + \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 \overset{\text{(d)}}{\geqslant} \frac{C-1}{C+1} \cdot \|\boldsymbol{\nu}\|_1.$$

Here, step (d) is a consequence of

$$\begin{aligned}
&(C{+}1) \cdot \big(- \|\boldsymbol{\nu}_{\mathcal{S}}\|_1 + \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 \big) \\
&\quad = -C \cdot \|\boldsymbol{\nu}_{\mathcal{S}}\|_1 - \|\boldsymbol{\nu}_{\mathcal{S}}\|_1 + C \cdot \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 + \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 \\
&\quad \overset{\text{(e)}}{\geqslant} -\|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 - \|\boldsymbol{\nu}_{\mathcal{S}}\|_1 + C \cdot \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 + C \cdot \|\boldsymbol{\nu}_{\mathcal{S}}\|_1 \\
&\quad = (C{-}1) \cdot \|\boldsymbol{\nu}_{\mathcal{S}}\|_1 + (C{-}1) \cdot \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 \\
&\quad = (C{-}1) \cdot \|\boldsymbol{\nu}\|_1,
\end{aligned}$$

where step (e) follows from applying twice the fact that $\boldsymbol{\nu} \in \mathrm{nullspace}_{\mathbb{R}}(\boldsymbol{H}_{\mathrm{CS}})$ and the assumption that $\boldsymbol{H}_{\mathrm{CS}} \in \mathrm{NSP}^{\leqslant}_{\mathbb{R}}(k, C)$. Subtracting the term $\|\boldsymbol{e}_{\mathcal{S}}\|_1$ on both sides of (2), and solving for $\|\boldsymbol{\nu}\|_1 = \|\boldsymbol{e} - \hat{\boldsymbol{e}}\|_1$ yields the promised result. ∎

IV. Channel Coding Linear Programming Decoding

A. The Setup

We consider coded data transmission over a memoryless channel with input alphabet $\mathcal{X}_{\mathrm{CC}} \triangleq \{0, 1\}$, output alphabet $\mathcal{Y}_{\mathrm{CC}}$, and channel law $P_{Y|X}(y|x)$ with the help of a binary linear code $\mathcal{C}_{\mathrm{CC}}$ of length n and dimension κ with $n \geqslant \kappa$. In the following, we will identify $\mathcal{X}_{\mathrm{CC}}$ with $\mathbb{F}_2$.

- Let $\boldsymbol{G}_{\mathrm{CC}} \in \mathbb{F}_2^{n \times \kappa}$ be a generator matrix for $\mathcal{C}_{\mathrm{CC}}$. Consequently, $\boldsymbol{G}_{\mathrm{CC}}$ has rank κ over $\mathbb{F}_2$, and information vectors $\boldsymbol{u} \in \mathbb{F}_2^{\kappa}$ are encoded into codewords $\boldsymbol{x} \in \mathbb{F}_2^n$ according to $\boldsymbol{x} = \boldsymbol{G}_{\mathrm{CC}} \cdot \boldsymbol{u}$ (in $\mathbb{F}_2$), i.e.,. $\mathcal{C}_{\mathrm{CC}} = \big\{ \boldsymbol{G}_{\mathrm{CC}} \cdot \boldsymbol{u} \text{ (in } \mathbb{F}_2) \,\big|\, \boldsymbol{u} \in \mathbb{F}_2^{\kappa} \big\}$.[3]
- Let $\boldsymbol{H}_{\mathrm{CC}} \in \mathbb{F}_2^{m \times n}$ be a parity-check matrix for $\mathcal{C}_{\mathrm{CC}}$. Consequently, $\boldsymbol{H}_{\mathrm{CC}}$ has rank $n - \kappa \leqslant m$ over $\mathbb{F}_2$, and any $\boldsymbol{x} \in \mathbb{F}_2^n$ satisfies $\boldsymbol{H}_{\mathrm{CC}} \cdot \boldsymbol{x} = \boldsymbol{0}$ (in $\mathbb{F}_2$) if and only if $\boldsymbol{x} \in \mathcal{C}_{\mathrm{CC}}$, i.e., $\mathcal{C}_{\mathrm{CC}} = \big\{ \boldsymbol{x} \in \mathbb{F}_2^n \,\big|\, \boldsymbol{H}_{\mathrm{CC}} \cdot \boldsymbol{x} = \boldsymbol{0} \text{ (in } \mathbb{F}_2) \big\}$.
- Let $\boldsymbol{y} \in \mathcal{Y}_{\mathrm{CC}}^n$ be the received vector and define for each $i \in \mathcal{I}(\boldsymbol{H}_{\mathrm{CC}})$ the log-likelihood ratio $\lambda_i \triangleq \lambda_i(y_i) \triangleq \log\big(\frac{P_{Y|X}(y_i|0)}{P_{Y|X}(y_i|1)}\big)$.
- On the side, let us remark that if $\mathcal{Y}_{\mathrm{CC}}$ is binary then $\mathcal{Y}_{\mathrm{CC}}$ can be identified with $\mathbb{F}_2$ and we can write $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{e}$ (in $\mathbb{F}_2$) for a suitably defined vector $\boldsymbol{e} \in \mathbb{F}_2^n$, which will be called the error vector. Moreover, we can define the syndrome vector $\boldsymbol{s} \triangleq \boldsymbol{H}_{\mathrm{CC}} \cdot \boldsymbol{y}$ (in $\mathbb{F}_2$). Note that

  $$\begin{aligned}
  \boldsymbol{s} &= \boldsymbol{H}_{\mathrm{CC}} \cdot (\boldsymbol{x} + \boldsymbol{e}) = \boldsymbol{H}_{\mathrm{CC}} \cdot \boldsymbol{x} + \boldsymbol{H}_{\mathrm{CC}} \cdot \boldsymbol{e} \\
  &= \boldsymbol{H}_{\mathrm{CC}} \cdot \boldsymbol{e} \quad \text{(in } \mathbb{F}_2).
  \end{aligned}$$

  However, in the following we will only use the log-likelihood ratio vector $\boldsymbol{\lambda}$ (that can be defined for any alphabet $\mathcal{Y}_{\mathrm{CC}}$), and not the binary syndrome vector $\boldsymbol{s}$.

Upon observing $\boldsymbol{Y} = \boldsymbol{y}$, the maximum-likelihood decoding (MLD) rule decides for $\hat{\boldsymbol{x}}(\boldsymbol{y}) = \arg\max_{\boldsymbol{x}' \in \mathcal{C}_{\mathrm{CC}}} P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}')$ where $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}') = \prod_{i \in \mathcal{I}} P_{Y|X}(y_i|x_i')$.[4] Formally:

CC-MLD1 :	maximize	$P_{\boldsymbol{Y}\|\boldsymbol{X}}(\boldsymbol{y}\|\boldsymbol{x}')$
	subject to	$\boldsymbol{x}' \in \mathcal{C}_{\mathrm{CC}}$.

It is clear that instead of $P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}')$ we can also maximize $\log P_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}') = \sum_{i \in \mathcal{I}} \log P_{Y|X}(y_i|x_i')$. Noting that $\log P_{Y|X}(y_i|x_i') = -\lambda_i x_i' + \log P_{Y|X}(y_i|0)$ for $x_i' \in \{0, 1\}$, **CC-MLD1** can then be rewritten to read

CC-MLD2 :	minimize	$\langle \boldsymbol{\lambda}, \boldsymbol{x}' \rangle$
	subject to	$\boldsymbol{x}' \in \mathcal{C}_{\mathrm{CC}}$.

Because the cost function is linear, and a linear function attains its minimum at the extremal points of a convex set, this is essentially equivalent to

CC-MLD3 :	minimize	$\langle \boldsymbol{\lambda}, \boldsymbol{x}' \rangle$
	subject to	$\boldsymbol{x}' \in \mathrm{conv}(\mathcal{C}_{\mathrm{CC}})$.

Although this is a linear program, it can usually not be solved efficiently because its description complexity is typically

[3] We remind the reader that throughout this paper we are using *column* vectors, which is in contrast to the coding theory habit to use *row* vectors.

[4] Actually, slightly more precise would be to call this decision rule "block-wise maximum-likelihood decoding."

exponential in the block length of the code.[5]

However, one might try to solve a relaxation of **CC-MLD3**. Namely, as proposed by Feldman, Wainwright, and Karger [3], [4], we can try to solve the optimization problem

$$\boxed{\begin{array}{lll} \textbf{CC-LPD}: & \text{minimize} & \langle \boldsymbol{\lambda}, \boldsymbol{x}' \rangle \\ & \text{subject to} & \boldsymbol{x}' \in \mathcal{P}(\boldsymbol{H}_{\mathrm{CC}}), \end{array}}$$

where the relaxed set $\mathcal{P}(\boldsymbol{H}_{\mathrm{CC}}) \supseteq \mathrm{conv}(C)$ is given in the next definition.

Definition 6 *For every $j \in \mathcal{J}(\boldsymbol{H}_{\mathrm{CC}})$, let $\boldsymbol{h}_j^{\mathsf{T}}$ be the j-th row of $\boldsymbol{H}_{\mathrm{CC}}$ and let $\mathcal{C}_{\mathrm{CC},j} \triangleq \left\{ \boldsymbol{x} \in \mathbb{F}_2^n \;\middle|\; \langle \boldsymbol{h}_j, \boldsymbol{x} \rangle = 0 \text{ (mod 2)} \right\}$. Then, the fundamental polytope $\mathcal{P} \triangleq \mathcal{P}(\boldsymbol{H}_{\mathrm{CC}})$ of $\boldsymbol{H}_{\mathrm{CC}}$ is defined to be the set*

$$\mathcal{P} \triangleq \mathcal{P}(\boldsymbol{H}_{\mathrm{CC}}) = \bigcap_{j \in \mathcal{J}} \mathrm{conv}(\mathcal{C}_{\mathrm{CC},j}).$$

Vectors in $\mathcal{P}(\boldsymbol{H}_{\mathrm{CC}})$ will be called pseudo-codewords. □

In order to motivate this relaxation, note that the code $\mathcal{C}$ can be written as

$$\mathcal{C}_{\mathrm{CC}} = \mathcal{C}_{\mathrm{CC},1} \cap \cdots \cap \mathcal{C}_{\mathrm{CC},m},$$

and so

$$\begin{aligned} \mathrm{conv}(\mathcal{C}_{\mathrm{CC}}) &= \mathrm{conv}(\mathcal{C}_{\mathrm{CC},1} \cap \cdots \cap \mathcal{C}_{\mathrm{CC},m}) \\ &\subseteq \mathrm{conv}(\mathcal{C}_{\mathrm{CC},1}) \cap \cdots \cap \mathrm{conv}(\mathcal{C}_{\mathrm{CC},m}) \\ &= \mathcal{P}(\boldsymbol{H}_{\mathrm{CC}}). \end{aligned}$$

It can be verified [3], [4] that this relaxation possesses the important property that all the vertices of $\mathrm{conv}(\mathcal{C}_{\mathrm{CC}})$ are also vertices of $\mathcal{P}(\boldsymbol{H}_{\mathrm{CC}})$. Let us emphasize that different parity-check matrices for the same code usually lead to different fundamental polytopes and therefore to different **CC-LPD**s.

Similarly to the compressed sensing setup, we want to understand when we can guarantee that the codeword estimate given by **CC-LPD** equals the codeword estimate given by **CC-MLD**. It is important to note, as we did in the compressed sensing setup, that we worry mostly about the solution given by **CC-LPD** being equal to the solution given by **CC-MLD**, because even **CC-MLD** might fail to correctly identify the codeword that was sent when the error vector is beyond the error correction capability of the code. Therefore, the performance of **CC-MLD** is a natural upper bound on the performance of **CC-LPD**, and a way to assess **CC-LPD** is to study the gap to **CC-MLD**, e.g., by comparing the performance guarantees for **CC-LPD** that are discussed here with known performance guarantees for **CC-MLD**.

When characterizing the **CC-LPD** performance of binary linear codes over binary-input output-symmetric channels [17] we can without loss of generality assume that the all-zero codeword was transmitted. With this, the success probability of **CC-LPD** is the probability that the all-zero codeword yields the lowest cost function value compared to all non-zero vectors in the fundamental polytope. Because the cost function is linear, this is equivalent to the statement that the success probability of **CC-LPD** equals the probability that the all-zero codeword yields the lowest cost function value compared to all non-zero vectors in the conic hull of the fundamental polytope. This conic hull is called the fundamental cone $\mathcal{K} \triangleq \mathcal{K}(\boldsymbol{H}_{\mathrm{CC}})$ and it can be written as

$$\mathcal{K} \triangleq \mathcal{K}(\boldsymbol{H}_{\mathrm{CC}}) = \mathrm{conic}\left(\mathcal{P}(\boldsymbol{H}_{\mathrm{CC}})\right) = \bigcap_{j \in \mathcal{J}} \mathrm{conic}(\mathcal{C}_{\mathrm{CC},j}).$$

The fundamental cone can be characterized by the inequalities listed in the following lemma [3], [4], [5], [6]. (Similar inequalities can be given for the fundamental polytope but we will not need them here.)

Lemma 7 *The fundamental cone $\mathcal{K} \triangleq \mathcal{K}(\boldsymbol{H}_{\mathrm{CC}})$ of $\boldsymbol{H}_{\mathrm{CC}}$ is the set of all vectors $\boldsymbol{\omega} \in \mathbb{R}^n$ that satisfy*

$$\omega_i \geqslant 0 \qquad \text{(for all } i \in \mathcal{I}\text{)}\,, \tag{3}$$

$$\omega_i \leqslant \sum_{i' \in \mathcal{I}_j \setminus i} \omega_{i'} \qquad \text{(for all } j \in \mathcal{J},\ \text{for all } i \in \mathcal{I}_j\text{)}\,. \tag{4}$$

A vector $\boldsymbol{\omega} \in \mathcal{K}$ is called a pseudo-codeword. If such a vector lies on an edge of $\mathcal{K}$, it is called a minimal pseudo-codeword. Moreover, if $\boldsymbol{\omega} \in \mathcal{K} \cap \mathbb{Z}^n$ and $\boldsymbol{\omega} \pmod 2 \in \mathcal{C}$, then $\boldsymbol{\omega}$ is called an unscaled pseudo-codeword. (For a motivation of these definitions, see [6], [18]).

Note that in the following, not only vectors in the fundamental polytope, but also vectors in the fundamental cone will be called pseudo-codewords. Moreover, if $\boldsymbol{H}_{\mathrm{CS}}$ is a zero-one measurement matrix, i.e., a measurement matrix where all entries are in $\{0, 1\}$, then we will consider $\boldsymbol{H}_{\mathrm{CS}}$ to represent also the parity-check matrix of some linear code over $\mathbb{F}_2$. Consequently, its fundamental polytope will be denoted by $\mathcal{P}(\boldsymbol{H}_{\mathrm{CS}})$ and its fundamental cone by $\mathcal{K}(\boldsymbol{H}_{\mathrm{CS}})$.

*B. Conditions for the Equivalence of **CC-LPD** and **CC-MLD***

The following lemma states when **CC-LPD** succeeds for the BSC.

Lemma 8 *Let $\boldsymbol{H}_{\mathrm{CC}}$ be the parity-check matrix of some code $\mathcal{C}_{\mathrm{CC}}$ and let $\mathcal{S} \subseteq \mathcal{I}(\boldsymbol{H}_{\mathrm{CC}})$ be the set of coordinate indices that are flipped by the BSC. If $\boldsymbol{H}_{\mathrm{CC}}$ is such that*

$$\|\boldsymbol{\omega}_{\mathcal{S}}\|_1 < \|\boldsymbol{\omega}_{\overline{\mathcal{S}}}\|_1 \tag{5}$$

*for all $\boldsymbol{\omega} \in \mathcal{K}(\boldsymbol{H}_{\mathrm{CC}}) \setminus \{\mathbf{0}\}$ then the **CC-LPD** decision equals the codeword that was sent.*

Remark: The above condition is also necessary, however, we will not use this fact in the following.

Proof: Without loss of generality, we can assume that the all-zero codeword was transmitted. Let $+L > 0$ be the log-likelihood ratio associated to a received 0, and let $-L < 0$ be the log-likelihood ratio associated to a received

[5]Examples of code families that have sub-exponential description complexities in the block length are convolutional codes (with fixed state-space size), cycle codes, and tree codes. However, these classes of codes are not good enough for achieving performance close to capacity even under ML decoding. (For more on this topic, see for example [16].)

1. Therefore, $\lambda_i = +L$ if $i \in \overline{\mathcal{S}}$ and $\lambda_i = -L$ if $i \in \mathcal{S}$. Then it follows from the assumptions in the lemma statement that for any $\boldsymbol{\omega} \in \mathcal{K}(\boldsymbol{H}_{\mathrm{CC}}) \setminus \{\mathbf{0}\}$

$$\begin{aligned}\langle \boldsymbol{\lambda}, \boldsymbol{\omega} \rangle &= \sum_{i \in \overline{\mathcal{S}}} (+L) \cdot \omega_i + \sum_{i \in \mathcal{S}} (-L) \cdot \omega_i \\ &\overset{\text{(a)}}{=} L \cdot \|\boldsymbol{\omega}_{\overline{\mathcal{S}}}\|_1 - L \cdot \|\boldsymbol{\omega}_{\mathcal{S}}\|_1 \overset{\text{(b)}}{>} 0 = \langle \boldsymbol{\lambda}, \mathbf{0} \rangle,\end{aligned}$$

where the equality follows from the fact that $|\omega_i| = \omega_i$ for all $i \in \mathcal{I}(\boldsymbol{H}_{\mathrm{CC}})$, and where the inequality in step (b) follows from (5). Therefore, under **CC-LPD** the all-zero codeword has the lowest cost function value compared to all the non-zero pseudo-codewords in the fundamental cone, and therefore also compared to all the non-zero pseudo-codewords in the fundamental polytope. ■

Note that the inequality in (5) is *identical* to the inequality that appears in the definition of the strict nullspace property for $C = 1$ (!) This observation makes one wonder if there is a connection between **CS-LPD** and **CC-LPD**, in particular for measurement matrices that contain only zeros and ones. Of course, in order to establish such a connection we first need to understand how points in the nullspace of the measurement matrix $\boldsymbol{H}_{\mathrm{CS}}$ can be associated with points in the fundamental polytope of the parity-check matrix $\boldsymbol{H}_{\mathrm{CS}}$ (now seen as a parity-check matrix for a code over $\mathbb{F}_2$). Such an association will be exhibited in Section V. However, before turning to that section, we will first discuss pseudo-weights, which are a popular way of characterizing the importance of the different pseudo-codewords in the fundamental cone and for establishing performance guarantees for **CC-LPD**.

C. Definition of Pseudo-Weights

Note that the fundamental polytope and cone are only a function of the parity-check matrix of the code and *not* of the channel. The influence of the channel is reflected in the pseudo-weight of the pseudo-codewords, so every channel has its pseudo-weight definition. Therefore, every communication channel comes with the right measure of distance that determines how often a fractional vertex is incorrectly chosen in **CC-LPD**.

Definition 9 ([19], [20], [3], [4], [5], [6]) *Let $\boldsymbol{\omega}$ be a non-zero vector in $\mathbb{R}^n_{\geqslant 0}$ with $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)$.*

- *The AWGNC (more precisely, binary-input AWGNC) pseudo-weight of $\boldsymbol{\omega}$ is defined to be*

$$w_{\mathrm{p}}^{\mathrm{AWGNC}}(\boldsymbol{\omega}) \triangleq \frac{\|\boldsymbol{\omega}\|_1^2}{\|\boldsymbol{\omega}\|_2^2}.$$

- *In order to define the BSC pseudo-weight $w_{\mathrm{p}}^{\mathrm{BSC}}(\boldsymbol{\omega})$, we let $\boldsymbol{\omega}'$ be the vector of length n with the same components as $\boldsymbol{\omega}$ but in non-increasing order. Now let*

$$\begin{aligned} f(\xi) &\triangleq \omega'_i \quad (i-1 < \xi \leqslant i,\ 0 < \xi \leqslant n), \\ F(\xi) &\triangleq \int_0^{\xi} f(\xi')\, \mathrm{d}\, \xi', \\ e &\triangleq F^{-1}\left(\frac{F(n)}{2}\right) = F^{-1}\left(\frac{\|\boldsymbol{\omega}\|_1}{2}\right). \end{aligned}$$

  Then the BSC pseudo-weight $w_{\mathrm{p}}^{\mathrm{BSC}}(\boldsymbol{\omega})$ of $\boldsymbol{\omega}$ is defined to be $w_{\mathrm{p}}^{\mathrm{BSC}}(\boldsymbol{\omega}) \triangleq 2e$.
- *The BEC pseudo-weight of $\boldsymbol{\omega}$ is defined to be*

$$w_{\mathrm{p}}^{\mathrm{BEC}}(\boldsymbol{\omega}) = \big|\operatorname{supp}(\boldsymbol{\omega})\big|.$$

- *The max-fractional weight of $\boldsymbol{\omega}$ is defined to be*

$$w_{\max-\mathrm{frac}}(\boldsymbol{\omega}) \triangleq \frac{\|\boldsymbol{\omega}\|_1}{\|\boldsymbol{\omega}\|_\infty}.$$

For $\boldsymbol{\omega} = \mathbf{0}$ we define all of the above pseudo-weights and the max-fractional weight to be zero. □

A detailed discussion of the motivation and significance of these definitions can be found in [6]. For a parity-check matrix $\boldsymbol{H}_{\mathrm{CC}}$ we define the minimum AWGNC pseudo-weight $w_{\mathrm{p}}^{\mathrm{AWGNC,min}}(\boldsymbol{H}_{\mathrm{CC}})$ to be

$$\begin{aligned} w_{\mathrm{p}}^{\mathrm{AWGNC,min}}(\boldsymbol{H}_{\mathrm{CC}}) &\triangleq \min_{\boldsymbol{\omega} \in \mathcal{P}(\boldsymbol{H}_{\mathrm{CC}}) \setminus \{\mathbf{0}\}} w_{\mathrm{p}}^{\mathrm{AWGNC}}(\boldsymbol{\omega}) \\ &= \min_{\boldsymbol{\omega} \in \mathcal{K}(\boldsymbol{H}_{\mathrm{CC}}) \setminus \{\mathbf{0}\}} w_{\mathrm{p}}^{\mathrm{AWGNC}}(\boldsymbol{\omega}). \end{aligned}$$

The minimum BSC pseudo-weight $w_{\mathrm{p}}^{\mathrm{BSC,min}}(\boldsymbol{H}_{\mathrm{CC}})$, the minimum BEC pseudo-weight $w_{\mathrm{p}}^{\mathrm{BEC,min}}(\boldsymbol{H}_{\mathrm{CC}})$, and the minimum max-fractional weight $w_{\max-\mathrm{frac}}^{\min}(\boldsymbol{H}_{\mathrm{CC}})$ of $\boldsymbol{H}_{\mathrm{CC}}$ are defined analogously. Note that although $w_{\max-\mathrm{frac}}^{\min}(\boldsymbol{H}_{\mathrm{CC}})$ yields weaker performance guarantees than the other quantities [6], it has the advantage of being efficiently computable [3], [4].

There are other possible definitions of a BSC pseudo-weight. For example, the BSC pseudo-weight of $\boldsymbol{\omega}$ can also be taken to be

$$w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega}) \triangleq \begin{cases} 2e & \text{if } \|\boldsymbol{\omega}'_{\{1,\ldots,e\}}\|_1 = \|\boldsymbol{\omega}'_{\{e+1,\ldots,n\}}\|_1 \\ 2e-1 & \text{if } \|\boldsymbol{\omega}'_{\{1,\ldots,e\}}\|_1 > \|\boldsymbol{\omega}'_{\{e+1,\ldots,n\}}\|_1 \end{cases},$$

where $\boldsymbol{\omega}'$ is defined as in Definition 9 and where e is the smallest integer such that $\|\boldsymbol{\omega}'_{\{1,\ldots,e\}}\|_1 \geqslant \|\boldsymbol{\omega}'_{\{e+1,\ldots,n\}}\|_1$. This definition of the BSC pseudo-weight was e.g. used in [21]. (Note that in [20] the quantity $w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega})$ was introduced as "BSC effective weight".)

Of course, the values $w_{\mathrm{p}}^{\mathrm{BSC}}(\boldsymbol{\omega})$ and $w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega})$ are tightly connected. Namely, if $w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega})$ is an even integer then $w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega}) = w_{\mathrm{p}}^{\mathrm{BSC}}(\boldsymbol{\omega})$, and if $w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega})$ is an odd integer then $w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega}) - 1 < w_{\mathrm{p}}^{\mathrm{BSC}}(\boldsymbol{\omega}) < w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega}) + 1$.

The following lemma establishes a connection between BSC pseudo-weights and the condition that appears in Lemma 8.

Lemma 10 *Let $\boldsymbol{H}_{\mathrm{CC}}$ be the parity-check matrix of some code $\mathcal{C}_{\mathrm{CC}}$ and let $\boldsymbol{\omega}$ be some arbitrary non-zero pseudo-codeword of $\boldsymbol{H}_{\mathrm{CC}}$, i.e., $\boldsymbol{\omega} \in \mathcal{K}(\boldsymbol{H}_{\mathrm{CC}}) \setminus \{\mathbf{0}\}$. Then for all sets $\mathcal{S} \subseteq \mathcal{I}$ with $\#\mathcal{S} < \frac{1}{2} \cdot w_{\mathrm{p}}^{\mathrm{BSC}}(\boldsymbol{\omega})$, or with $\#\mathcal{S} < \frac{1}{2} \cdot w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega})$, it holds that*

$$\|\boldsymbol{\omega}_{\mathcal{S}}\|_1 < \|\boldsymbol{\omega}_{\overline{\mathcal{S}}}\|_1.$$

Proof: First, consider the statement under for the assumption $\#\mathcal{S} < \frac{1}{2} \cdot w_{\mathrm{p}}^{\mathrm{BSC}}(\boldsymbol{\omega})$. The proof is by contradiction. So, assume that $\|\boldsymbol{\omega}_{\mathcal{S}}\|_1 \geqslant \|\boldsymbol{\omega}_{\overline{\mathcal{S}}}\|_1$ holds. This statement is

clearly equivalent to the statement that $2 \cdot \|\boldsymbol{\omega}_{\mathcal{S}}\|_1 \geqslant \|\boldsymbol{\omega}_{\mathcal{S}}\|_1 + \|\boldsymbol{\omega}_{\overline{\mathcal{S}}}\|_1 = \|\boldsymbol{\omega}\|_1$, which is equivalent to the statement that $\|\boldsymbol{\omega}_{\mathcal{S}}\|_1 \geqslant \frac{1}{2} \cdot \|\boldsymbol{\omega}\|_1$. In terms of the notation in Definition 9, this means that

$$w_{\mathrm{p}}^{\mathrm{BSC}}(\boldsymbol{\omega}) = 2 \cdot F^{-1}\left(\frac{\|\boldsymbol{\omega}\|_1}{2}\right) \overset{\text{(a)}}{\leqslant} 2 \cdot F^{-1}(\|\boldsymbol{\omega}_{\mathcal{S}}\|_1)$$
$$\overset{\text{(b)}}{\leqslant} 2 \cdot \frac{\|\boldsymbol{\omega}_{\mathcal{S}}\|_1}{\|\boldsymbol{\omega}\|_\infty} \leqslant 2 \cdot \frac{\#\mathcal{S} \cdot \|\boldsymbol{\omega}\|_\infty}{\|\boldsymbol{\omega}\|_\infty} = 2 \cdot \#\mathcal{S},$$

where at step (a) we have used the fact that F^{-1} is a (strictly) non-decreasing function and where at step (b) we have used the fact that the slope of F^{-1} (over the domain where F^{-1} is defined) is at least $1/\|\boldsymbol{\omega}\|_\infty$. This, however, is a contradiction to the assumption that $\#\mathcal{S} < \frac{1}{2} \cdot w_{\mathrm{p}}^{\mathrm{BSC}}(\boldsymbol{\omega})$.

Secondly, consider the statement under for the assumption $\#\mathcal{S} < \frac{1}{2} \cdot w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega})$. The proof is by contradiction. So, assume that the $\|\boldsymbol{\omega}_{\mathcal{S}}\|_1 \geqslant \|\boldsymbol{\omega}_{\overline{\mathcal{S}}}\|_1$ holds. With this, and the above definition of $\boldsymbol{\omega}'$ based on $\boldsymbol{\omega}$, $\|\boldsymbol{\omega}'_{\{1,\ldots,\#\mathcal{S}\}}\|_1 \geqslant \|\boldsymbol{\omega}_{\mathcal{S}}\|_1 \geqslant \|\boldsymbol{\omega}_{\overline{\mathcal{S}}}\|_1 \geqslant \|\boldsymbol{\omega}'_{\{\#\mathcal{S}+1,\ldots,n\}}\|_1$. If $w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega})$ is an even integer then this line of inequalities shows that $\#\mathcal{S} \geqslant \frac{1}{2} \cdot w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega})$, which is a contradiction to the assumption that $\#\mathcal{S} < \frac{1}{2} \cdot w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega})$. If $w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega})$ is an odd integer then this line of inequalities shows that $\#\mathcal{S} \geqslant \frac{1}{2} \cdot \left(w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega}) + 1\right) > \frac{1}{2} w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega})$, which again is a contradiction to the assumption that $\#\mathcal{S} < \frac{1}{2} \cdot w_{\mathrm{p}}^{\mathrm{BSC}'}(\boldsymbol{\omega})$. ■

V. Establishing a Bridge Between **CS-LPD** and **CC-LPD**

We are now ready to establish a bridge between **CS-LPD** and **CC-LPD**. Our main tool is a simple lemma that was already established in [22] but for a different purpose.

Lemma 11 *Let $\boldsymbol{H}_{\mathrm{CS}}$ be a measurement matrix that contains only zeros and ones. Then*

$$\boldsymbol{\nu} \in \mathrm{nullspace}_{\mathbb{R}}(\boldsymbol{H}_{\mathrm{CS}}) \quad \Rightarrow \quad |\boldsymbol{\nu}| \in \mathcal{K}(\boldsymbol{H}_{\mathrm{CS}}).$$

Remark: Note that $\mathrm{supp}(\boldsymbol{\nu}) = \mathrm{supp}(|\boldsymbol{\nu}|)$.

Proof: Let $\boldsymbol{\omega} \triangleq |\boldsymbol{\nu}|$. In order to show that such a vector $\boldsymbol{\omega}$ is indeed in the fundamental cone of $\boldsymbol{H}_{\mathrm{CS}}$, we need to verify (3) and (4). The way $\boldsymbol{\omega}$ is defined, it is clear that it satisfies (3). Therefore, let us focus on the proof that $\boldsymbol{\omega}$ satisfies (4). Namely, from $\boldsymbol{\nu} \in \mathrm{nullspace}_{\mathbb{R}}(\boldsymbol{H}_{\mathrm{CS}})$ it follows that for all $j \in \mathcal{J}$, $\sum_{i \in \mathcal{I}} h_{j,i} \nu_i = 0$, i.e., for all $j \in \mathcal{J}$, $\sum_{i \in \mathcal{I}_j} \nu_i = 0$. This implies

$$\omega_i = |\nu_i| = \left| - \sum_{i' \in \mathcal{I}_j \setminus i} \nu_{i'} \right| \leqslant \sum_{i' \in \mathcal{I}_j \setminus i} |\nu_{i'}| = \sum_{i' \in \mathcal{I}_j \setminus i} \omega_{i'}$$

for all $j \in \mathcal{J}$ and all $i \in \mathcal{I}_j$, showing that $\boldsymbol{\omega}$ indeed satisfies (4). ■

This lemma is fundamentally one-way: it says that with every point in the real nullspace of the measurement matrix $\boldsymbol{H}_{\mathrm{CS}}$ we can associate a point in the fundamental cone of $\boldsymbol{H}_{\mathrm{CS}}$, but not necessarily vice-versa. Therefore a problematic point for the real nullspace of $\boldsymbol{H}_{\mathrm{CS}}$ will translate to a problematic point in the fundamental cone of $\boldsymbol{H}_{\mathrm{CS}}$ and hence to bad performance of **CC-LPD**. Similarly, a "good" parity-check matrix $\boldsymbol{H}_{\mathrm{CS}}$ must have no low pseudo-weight points in the fundamental cone, which means that there are no problematic points in the real nullspace of $\boldsymbol{H}_{\mathrm{CS}}$. Therefore "positive" results for channel coding will translate into "positive" results for compressed sensing, and "negative" results for compressed sensing will translate into "negative" results for channel coding.

Further, the lemma preserves the support of a given point $\boldsymbol{\nu}$. That means that if there are no low pseudo-weight points in the fundamental cone of $\boldsymbol{H}_{\mathrm{CS}}$ with a given support, there are no problematic points in the real nullspace of $\boldsymbol{H}_{\mathrm{CS}}$ with the same support, which allows point-wise versions of all our results.

VI. Translation of Performance Guarantees

In this section we use the bridge between **CS-LPD** and **CC-LPD** that was established in the previous section to translate "positive" results about **CC-LPD** to "positive" results about **CS-LPD**.

A. The Role of the BSC Pseudo-Weight for **CS-LPD**

Lemma 12 *Let $\boldsymbol{H}_{\mathrm{CS}} \in \{0,1\}^{m \times n}$ be a CS measurement matrix and let k be a non-negative integer. Then*

$$w_{\mathrm{p}}^{\mathrm{BSC,min}}(\boldsymbol{H}_{\mathrm{CS}}) > 2k \quad \Rightarrow \quad \boldsymbol{H}_{\mathrm{CS}} \in \mathrm{NSP}_{\mathbb{R}}^{<}(k, C\!=\!1).$$

Proof: Fix some $\boldsymbol{\nu} \in \mathrm{nullspace}_{\mathbb{R}}(\boldsymbol{H}_{\mathrm{CS}}) \setminus \{\mathbf{0}\}$. By Lemma 11 we know that $|\boldsymbol{\nu}|$ is a pseudo-codeword of $\boldsymbol{H}_{\mathrm{CS}}$, and by the assumption $w_{\mathrm{p}}^{\mathrm{BSC,min}}(\boldsymbol{H}_{\mathrm{CS}}) > 2k$ we know that $w_{\mathrm{p}}^{\mathrm{BSC}}(|\boldsymbol{\nu}|) > 2k$. Then, using Lemma 10, we conclude that for all sets $\mathcal{S} \subseteq \mathcal{I}$ with $\#\mathcal{S} \leqslant k$, we must have $\|\boldsymbol{\nu}_{\mathcal{S}}\|_1 = \|\,|\boldsymbol{\nu}_{\mathcal{S}}|\,\|_1 < \|\,|\boldsymbol{\nu}_{\overline{\mathcal{S}}}|\,\|_1 = \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1$. Because $\boldsymbol{\nu}$ was arbitrary, the claim $\boldsymbol{H}_{\mathrm{CS}} \in \mathrm{NSP}_{\mathbb{R}}^{<}(k, C\!=\!1)$ clearly follows. ■

Recent results on the performance analysis of **CC-LPD** showed that parity-check matrices constructed from expander graphs can correct a constant fraction (of the block length n) of worst case [23] and random [8], [24] errors. (These types of results are analogous to the so-called strong and weak bounds for compressed sensing, respectively.)

These worst case error performance guarantees implicitly show that the BSC pseudo-weight of all pseudo-codewords of a binary linear code defined by a Tanner with sufficient expansion (strictly larger than $3/4$) must grow linearly in n. (A conclusion in a similar direction can be drawn for the random error setup.) We can therefore use our results to obtain new performance guarantees for **CS-LPD** based sparse recovery problems.

Let us mention that in [9], [25] expansion arguments were used to directly obtain similar types of performance guarantees for compressed sensing; the comparison of these guarantees to the guarantees that can be obtained through our channel-coding-based arguments remains as future work.

B. The Role of Binary-Input Channels Beyond the BSC for *CS-LPD*

In Lemma 12 we made a connection between performance guarantees for the BSC under **CC-LPD** on the one hand and the strict nullspace property $\mathrm{NSP}^{<}_{\mathbb{R}}(k, C)$ for $C = 1$ on the other hand. In this subsection we want to mention that one can establish a connection between performance guarantees for a certain class of binary-input channels under **CS-LPD** and the strict nullspace property $\mathrm{NSP}^{<}_{\mathbb{R}}(k, C)$ for $C > 1$. This class of channels consists of binary-input memoryless channels where for all output symbols the magnitude of the log-likelihood ratio is bounded by some constant $W \in \mathbb{R}_{>0}$. Without going into the details, the results from [26] (which generalize results from [23]) can be used to establish this connection.[6]

The results of this section will be discussed in more detail in a longer version of the present paper.

C. Connection between AWGNC Pseudo-Weight and ℓ_2/ℓ_1 Guarantees

Theorem 13 *Let $\boldsymbol{H}_{\mathrm{CS}} \in \{0,1\}^{m\times n}$ be a measurement matrix and let $\boldsymbol{s}$ and $\boldsymbol{e}$ be such that $\boldsymbol{s} = \boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{e}$. Moreover, let $\mathcal{S} \subseteq \mathcal{I}(\boldsymbol{H}_{\mathrm{CS}})$ with $\#\mathcal{S} = k$, and let C' be an arbitrary positive real number with $C' > 4k$. Then the estimate $\hat{\boldsymbol{e}}$ produced by **CS-LPD** will satisfy*

$$\|\boldsymbol{e} - \hat{\boldsymbol{e}}\|_2 \leqslant \frac{C''}{\sqrt{k}} \cdot \|\boldsymbol{e}_{\overline{\mathcal{S}}}\|_1 \qquad \text{with} \qquad C'' \triangleq \frac{1}{\sqrt{\frac{C'}{4k}} - 1},$$

if $w_{\mathrm{p}}^{\mathrm{AWGNC}}(|\boldsymbol{\nu}|) \geqslant C'$ holds for all $\boldsymbol{\nu} \in \mathrm{nullspace}_{\mathbb{R}}(\boldsymbol{H}_{\mathrm{CS}}) \setminus \{\boldsymbol{0}\}$. (In particular, this latter condition is satisfied for a measurement matrix $\boldsymbol{H}_{\mathrm{CS}}$ with $w_{\mathrm{p}}^{\mathrm{AWGNC,min}}(\boldsymbol{H}_{\mathrm{CS}}) \geqslant C'$.)

Proof: By definition, $\boldsymbol{e}$ is the original signal. Since $\boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{e} = \boldsymbol{s}$ and $\boldsymbol{H}_{\mathrm{CS}} \cdot \hat{\boldsymbol{e}} = \boldsymbol{s}$, it easily follows that $\boldsymbol{\nu} \triangleq \boldsymbol{e} - \hat{\boldsymbol{e}}$ is in the nullspace of $\boldsymbol{H}_{\mathrm{CS}}$. So,

$$\begin{aligned}
\|\boldsymbol{e}_{\mathcal{S}}\|_1 + \|\boldsymbol{e}_{\overline{\mathcal{S}}}\|_1 &= \|\boldsymbol{e}\|_1 \\
&\overset{\text{(a)}}{\geqslant} \|\hat{\boldsymbol{e}}\|_1 \\
&= \|\boldsymbol{e} + \boldsymbol{\nu}\|_1 \\
&= \|\boldsymbol{e}_{\mathcal{S}} + \boldsymbol{\nu}_{\mathcal{S}}\|_1 + \|\boldsymbol{e}_{\overline{\mathcal{S}}} + \boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 \\
&\overset{\text{(b)}}{\geqslant} \|\boldsymbol{e}_{\mathcal{S}}\|_1 - \|\boldsymbol{\nu}_{\mathcal{S}}\|_1 + \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 - \|\boldsymbol{e}_{\overline{\mathcal{S}}}\|_1 \\
&\overset{\text{(c)}}{\geqslant} \|\boldsymbol{e}_{\mathcal{S}}\|_1 + \left(\sqrt{C'} - 2\sqrt{k}\right)\|\boldsymbol{\nu}\|_2 - \|\boldsymbol{e}_{\overline{\mathcal{S}}}\|_1, \quad (6)
\end{aligned}$$

where step (a) follows from the fact that the solution to **CS-LPD** satisfies $\|\hat{\boldsymbol{e}}\|_1 \leqslant \|\boldsymbol{e}\|_1$ and where step (b) follows from applying the triangle inequality for the ℓ_1 norm twice. Moreover, step (c) follows from

$$\begin{aligned}
-\|\boldsymbol{\nu}_{\mathcal{S}}\|_1 + \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 &= \|\boldsymbol{\nu}\|_1 - 2\|\boldsymbol{\nu}_{\mathcal{S}}\|_1 \\
&\overset{\text{(d)}}{\geqslant} \sqrt{C'}\|\boldsymbol{\nu}\|_2 - 2\|\boldsymbol{\nu}_{\mathcal{S}}\|_1 \\
&\overset{\text{(e)}}{\geqslant} \sqrt{C'}\|\boldsymbol{\nu}\|_2 - 2\sqrt{k}\|\boldsymbol{\nu}_{\mathcal{S}}\|_2 \\
&\overset{\text{(f)}}{\geqslant} \sqrt{C'}\|\boldsymbol{\nu}\|_2 - 2\sqrt{k}\|\boldsymbol{\nu}\|_2 \\
&= \left(\sqrt{C'} - 2\sqrt{k}\right)\|\boldsymbol{\nu}\|_2,
\end{aligned}$$

where step (d) follows from the assumption that $w_{\mathrm{p}}^{\mathrm{AWGNC}}(|\boldsymbol{\nu}|) \geqslant C'$ for all $\boldsymbol{\nu} \in \mathrm{nullspace}_{\mathbb{R}}(\boldsymbol{H}_{\mathrm{CS}}) \setminus \{\boldsymbol{0}\}$, i.e., $\|\boldsymbol{\nu}\|_1 \geqslant \sqrt{C'} \cdot \|\boldsymbol{\nu}\|_2$ for all $\boldsymbol{\nu} \in \mathrm{nullspace}_{\mathbb{R}}(\boldsymbol{H}_{\mathrm{CS}})$, where step (e) follows from the inequality $\|\boldsymbol{a}\|_1 \leqslant \sqrt{k} \cdot \|\boldsymbol{a}\|_2$ that holds for any real vector $\boldsymbol{a}$ of length k, and where step (f) follows the inequality $\|\boldsymbol{a}_{\mathcal{S}}\|_2 \leqslant \|\boldsymbol{a}\|_2$ that holds for any real vector $\boldsymbol{a}$ whose set of coordinate indices includes $\mathcal{S}$. Subtracting the term $\|\boldsymbol{e}_{\mathcal{S}}\|_1$ on both sides of (6), and solving for $\|\boldsymbol{\nu}\|_2 = \|\boldsymbol{e} - \hat{\boldsymbol{e}}\|_2$ yields the promised result. ■

D. Connection between Max-Fractional Weight and ℓ_∞/ℓ_1 Guarantees

Theorem 14 *Let $\boldsymbol{H}_{\mathrm{CS}} \in \{0,1\}^{m\times n}$ be a measurement matrix and let $\boldsymbol{s}$ and $\boldsymbol{e}$ be such that $\boldsymbol{s} = \boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{e}$. Moreover, let $\mathcal{S} \subseteq \mathcal{I}(\boldsymbol{H}_{\mathrm{CS}})$ with $\#\mathcal{S} = k$, and let C' be an arbitrary positive real number with $C' > 2k$. Then the estimate $\hat{\boldsymbol{e}}$ produced by **CS-LPD** will satisfy*

$$\|\boldsymbol{e} - \hat{\boldsymbol{e}}\|_\infty \leqslant \frac{C''}{k} \cdot \|\boldsymbol{e}_{\overline{\mathcal{S}}}\|_1 \qquad \text{with} \qquad C'' \triangleq \frac{1}{\frac{C'}{2k} - 1},$$

if $w_{\mathrm{max-frac}}(|\boldsymbol{\nu}|) \geqslant C'$ holds for all $\boldsymbol{\nu} \in \mathrm{nullspace}_{\mathbb{R}}(\boldsymbol{H}_{\mathrm{CS}}) \setminus \{\boldsymbol{0}\}$. (In particular, this latter condition is satisfied for a measurement matrix $\boldsymbol{H}_{\mathrm{CS}}$ with $w_{\mathrm{max-frac}}^{\mathrm{min}}(\boldsymbol{H}_{\mathrm{CS}}) \geqslant C'$.)

Proof: By definition, $\boldsymbol{e}$ is the original signal. Since $\boldsymbol{H}_{\mathrm{CS}} \cdot \boldsymbol{e} = \boldsymbol{s}$ and $\boldsymbol{H}_{\mathrm{CS}} \cdot \hat{\boldsymbol{e}} = \boldsymbol{s}$, it easily follows that $\boldsymbol{\nu} \triangleq \boldsymbol{e} - \hat{\boldsymbol{e}}$ is in the nullspace of $\boldsymbol{H}_{\mathrm{CS}}$. So,

$$\begin{aligned}
\|\boldsymbol{e}_{\mathcal{S}}\|_1 + \|\boldsymbol{e}_{\overline{\mathcal{S}}}\|_1 &= \|\boldsymbol{e}\|_1 \\
&\overset{\text{(a)}}{\geqslant} \|\hat{\boldsymbol{e}}\|_1 \\
&= \|\boldsymbol{e} + \boldsymbol{\nu}\|_1 \\
&= \|\boldsymbol{e}_{\mathcal{S}} + \boldsymbol{\nu}_{\mathcal{S}}\|_1 + \|\boldsymbol{e}_{\overline{\mathcal{S}}} + \boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 \\
&\overset{\text{(b)}}{\geqslant} \|\boldsymbol{e}_{\mathcal{S}}\|_1 - \|\boldsymbol{\nu}_{\mathcal{S}}\|_1 + \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 - \|\boldsymbol{e}_{\overline{\mathcal{S}}}\|_1 \\
&\overset{\text{(c)}}{\geqslant} \|\boldsymbol{e}_{\mathcal{S}}\|_1 + (C' - 2k) \cdot \|\boldsymbol{\nu}\|_\infty - \|\boldsymbol{e}_{\overline{\mathcal{S}}}\|_1, \quad (7)
\end{aligned}$$

where step (a) follows from the fact that the solution to **CS-LPD** satisfies $\|\hat{\boldsymbol{e}}\|_1 \leqslant \|\boldsymbol{e}\|_1$ and where step (b) follows from applying the triangle inequality for the ℓ_1 norm twice.

[6] Note that in [26], "This suggests that the asymptotic advantage over [...] is gained not by quantization, but rather by restricting the LLRs to have finite support." should read "This suggests that the asymptotic advantage over [...] is gained not by quantization, but rather by restricting the LLRs to have bounded support."

Moreover, step (c) follows from

$$\begin{aligned}-\|\boldsymbol{\nu}_{\mathcal{S}}\|_1 + \|\boldsymbol{\nu}_{\overline{\mathcal{S}}}\|_1 &= \|\boldsymbol{\nu}\|_1 - 2 \cdot \|\boldsymbol{\nu}_{\mathcal{S}}\|_1 \\ &\overset{\text{(d)}}{\geqslant} C' \cdot \|\boldsymbol{\nu}\|_\infty - 2 \cdot \|\boldsymbol{\nu}_{\mathcal{S}}\|_1 \\ &\overset{\text{(e)}}{\geqslant} C' \cdot \|\boldsymbol{\nu}\|_\infty - 2k \cdot \|\boldsymbol{\nu}_{\mathcal{S}}\|_\infty \\ &\overset{\text{(f)}}{\geqslant} \sqrt{C'} \cdot \|\boldsymbol{\nu}\|_\infty - 2k \cdot \|\boldsymbol{\nu}\|_\infty \\ &= (C' - 2k) \cdot \|\boldsymbol{\nu}\|_\infty,\end{aligned}$$

where step (d) follows from the assumption that $w_{\max-\text{frac}}(|\boldsymbol{\nu}|) \geqslant C'$ for all $\boldsymbol{\nu} \in \text{nullspace}_{\mathbb{R}}(\boldsymbol{H}_{\text{CS}}) \setminus \{\boldsymbol{0}\}$, i.e., $\|\boldsymbol{\nu}\|_1 \geqslant C' \cdot \|\boldsymbol{\nu}\|_\infty$ for all $\boldsymbol{\nu} \in \text{nullspace}_{\mathbb{R}}(\boldsymbol{H}_{\text{CS}})$, where step (e) follows from the inequality $\|\boldsymbol{a}\|_1 \leqslant k \cdot \|\boldsymbol{a}\|_\infty$ that holds for any real vector $\boldsymbol{a}$ of length k, and where step (f) follows the inequality $\|\boldsymbol{a}_{\mathcal{S}}\|_\infty \leqslant \|\boldsymbol{a}\|_\infty$ that holds for any real vector $\boldsymbol{a}$ whose set of coordinate indices includes $\mathcal{S}$. Subtracting the term $\|\boldsymbol{e}_{\mathcal{S}}\|_1$ on both sides of (7), and solving for $\|\boldsymbol{\nu}\|_\infty = \|\boldsymbol{e} - \hat{\boldsymbol{e}}\|_\infty$ yields the promised result. ■

E. Connection between BEC Pseudo-Weight and ***CS-LPD***

For the binary erasure channel, **CC-LPD** is identical to the peeling decoder [17] that is just solving a system of linear equations by only using back-substitution. We can define an analogous compressed sensing problem by assuming that the compressed sensing decoder is *given the support* of the sparse signal $\boldsymbol{e}$ and decoding simply involves trying to recover the values of the non-zero entries by back-substitution, similarly to iterative matching pursuit. In this case it is clear that **CC-LPD** for the BEC and the described compressed sensing decoder have identical performance since back-substitution behaves exactly the same way over any field, be it the field of real numbers or any finite field. (Note that whereas the result of the **CC-LPD** for the BEC equals the result of the back-substitution-based decoder for the BEC, the same is not true for compressed sensing, i.e., **CS-LPD** with given support of the sparse signal can be strictly better than the back-substitution-based decoder with given support of the sparse signal.)

VII. Conclusions and Future work

Based on the observation that points in the nullspace of a zero-one matrix (considered as a real measurement matrix) can be mapped to points in the fundamental cone of the same matrix (considered as the parity-check matrix of a code over $\mathbb{F}_2$), we were able to establish a connection between **CS-LPD** and **CC-LPD**.

In addition to **CS-LPD**, a number of combinatorial algorithms (e.g. [27], [25], [28], [9], [29]) have been proposed for compressed sensing problems, with the benefit of faster decoding complexity and comparable performance to **CS-LPD**. It would be interesting to investigate if the connection of sparse recovery problems to channel coding extends in a similar manner for these decoders. One example of such a clear connection is the bit-flipping algorithm of Sipser and Spielman [30] and the corresponding algorithm for compressed sensing by Xu and Hassibi [25]. Connections of message-passing decoders for compressed sensing problems were also recently discussed in [31].

Other interesting directions involve using optimized channel coding matrices with randomized or deterministic constructions (e.g., see [17]) to create measurement matrices. Another is using ideas for improving the performance of a given measurement matrix (for example by removing short cycles), with possible theoretical guarantees. Finally, one interesting question relates to being able to certify in polynomial time that a given measurement matrix has good performance.

In any case, we hope that the connection between **CS-LPD** and **CC-LPD** that was discussed in this paper will help deepen the understanding of the role of linear programming relaxations for sparse recovery and for channel coding, in particular by translating results from one field to the other.

Acknowledgments

The first author would like to thank Prof. Babak Hassibi for stimulating discussions that helped in the development of this research.

References

[1] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] J. Feldman, *Decoding Error-Correcting Codes via Linear Programming*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 2003.

[4] J. Feldman, M. J. Wainwright, and D. R. Karger, "Using linear programming to decode binary linear codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 954–972, Mar. 2005.

[5] R. Koetter and P. O. Vontobel, "Graph covers and iterative decoding of finite-length codes," in *Proc. 3rd Intern. Symp. on Turbo Codes and Related Topics*, Brest, France, Sept. 1–5 2003, pp. 75–82.

[6] P. O. Vontobel and R. Koetter, "Graph-cover decoding and finite-length analysis of message-passing iterative decoding of LDPC codes," *accepted for IEEE Trans. Inform. Theory, available online under* `http://www.arxiv.org/abs/cs.IT/0512078`, 2007.

[7] J. Feldman, T. Malkin, R. A. Servedio, C. Stein, and M. J. Wainwright, "LP decoding corrects a constant fraction of errors," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, USA, June 27–July 2 2004, p. 68.

[8] C. Daskalakis, A. G. Dimakis, R. M. Karp, and M. J. Wainwright, "Probabilistic analysis of linear programming decoding," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3565–3578, Aug. 2008.

[9] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, "Combining geometry and combinatorics: a unified approach to sparse signal recovery," in *Proc. 46th Allerton Conf. on Communications, Control, and Computing*, Allerton House, Monticello, Illinois, USA, Sept. 23–26 2008.

[10] W. Xu and B. Hassibi, "Compressed sensing over the Grassmann manifold: A unified analytical framework," *preprint*, 2008.

[11] M. Stojnic, W. Xu, and B. Hassibi, "Compressed sensing - probabilistic analysis of a null-space characterization," in *Proc. IEEE Intern. Conf. Acoustics, Speech and Signal Processing*, Las Vegas, NV, USA, Mar. 31–Apr. 4 2008, pp. 3377–3380.

[12] Y. Zhang, "A simple proof for recoverability of ℓ_1-minimization: go over or under?," *Rice CAAM Department Technical Report TR05-09*, 2005.

[13] N. Linial and I. Novik, "How neighborly can a centrally symmetric polytope be?," *J. Discr. and Comp. Geom.*, vol. 36, no. 2, pp. 273–281, Sept. 2006.

[14] A. Feuer and A. Nemirovski, "On sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1579–1581, June 2003.

[15] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *J. Amer. Math. Soc.*, vol. 22, pp. 211–231, July 2008.

[16] N. Kashyap, "A decomposition theory for binary linear codes," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3035–3058, July 2008.

[17] T. Richardson and R. Urbanke, *Modern Coding Theory*, Cambridge University Press, New York, NY, 2008.

[18] R. Koetter, W.-C. W. Li, P. O. Vontobel, and J. L. Walker, "Characterizations of pseudo-codewords of (low-density) parity-check codes," *Adv. in Math.*, vol. 213, no. 1, pp. 205–229, Aug. 2007.

[19] N. Wiberg, *Codes and Decoding on General Graphs*, Ph.D. thesis, Linköping University, Sweden, 1996.

[20] G. D. Forney, Jr., R. Koetter, F. R. Kschischang, and A. Reznik, "On the effective weights of pseudocodewords for codes defined on graphs with cycles," in *Codes, Systems, and Graphical Models (Minneapolis, MN, 1999)*, B. Marcus and J. Rosenthal, Eds., vol. 123 of *IMA Vol. Math. Appl.*, pp. 101–112. Springer Verlag, New York, Inc., 2001.

[21] C. A. Kelley and D. Sridhara, "Pseudocodewords of Tanner graphs," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4013–4038, Nov. 2007.

[22] R. Smarandache and P. O. Vontobel, "Absdet-pseudo-codewords and perm-pseudo-codewords: definitions and properties," in *Proc. IEEE Int. Symp. Information Theory*, Seoul, Korea, June 28–July 3 2009.

[23] J. Feldman, T. Malkin, R. A. Servedio, C. Stein, and M. J. Wainwright, "LP decoding corrects a constant fraction of errors," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 82–89, Jan. 2007.

[24] S. Arora, C. Daskalakis, and D. Steurer, "Message-passing algorithms and improved LP decoding," in *Proc. of the 38th Annual ACM Symp. Theory of Computing*, Bethesda, MD, USA, May 31–June 2 2009.

[25] W. Xu and B. Hassibi, "Efficient compressive sensing with deterministic guarantees using expander graphs," in *Proc. IEEE Information Theory Workshop*, Tahoe City, CA, USA, Sept. 2–6 2007, pp. 414–419.

[26] J. Feldman, R. Koetter, and P. O. Vontobel, "The benefit of thresholding in LP decoding of LDPC codes," in *Proc. IEEE Int. Symp. Information Theory*, Adelaide, Australia, Sep. 4–9 2005, pp. 307–311.

[27] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing: closing the gap between performance and complexity," *submitted, available online under* `http://arxiv.org/abs/0803.0811`, Mar. 2008.

[28] J. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[29] V. Guruswami, J. Lee, and A. Wigderson, "Euclidean sections with sublinear randomness and error-correction over the reals," in *Proc. 12th Intern. Workshop on Randomization and Computation*, Cambridge, MA, USA, Aug. 25–27 2008.

[30] M. Sipser and D. Spielman, "Expander codes," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1710–1722, Nov. 1996.

[31] F. Zhang and H. D. Pfister, "On the iterative decoding of high rate LDPC codes with applications in compressed sensing," *submitted, available online under* `http://arxiv.org/abs/0903.2232`, Mar. 2009.